- Main objectives in time series analysis:
  - to understand the underlying structure of the time series by breaking it down to its components (trend - seasonality - cycle - random
  - Fit a mathematical model and then proceed to forecast future
- Time Series Analysis Components:
  - Trend Component: Trend is along term movement in a time series. It's the underlying direction (upward or downward) that can be ⊕ or ⊖ depending on whether the time series exhibits an increasing or decreasing long term pattern
  - Seasonal Component: Component of variation in a time series which is dependent on time of the year, describes any regular variation with a period of less than one year
  - Cyclic Component: Cyclical variations of non-seasonal nature whose periodicity is unknown
  - Random Component: - Random values left over when other components (trend - seasonal - cyclic) have been accounted for

Reminder: we use time series analysis to forecast future events

Method: Box-Jenkins (ARMA)

- use cases: forecast (next month sales - tomorrow's stock price)
- Time series data provide useful information about systems generating the time series such as:
  - Economics/Finance: prices, profits, imports, exports
  - Sociology: school enrollments, crime rate
  - Environment: Amount of pollutants
  - Meterology: Rainfall, temperature
  - Medicine: Blood pressure measurements over time

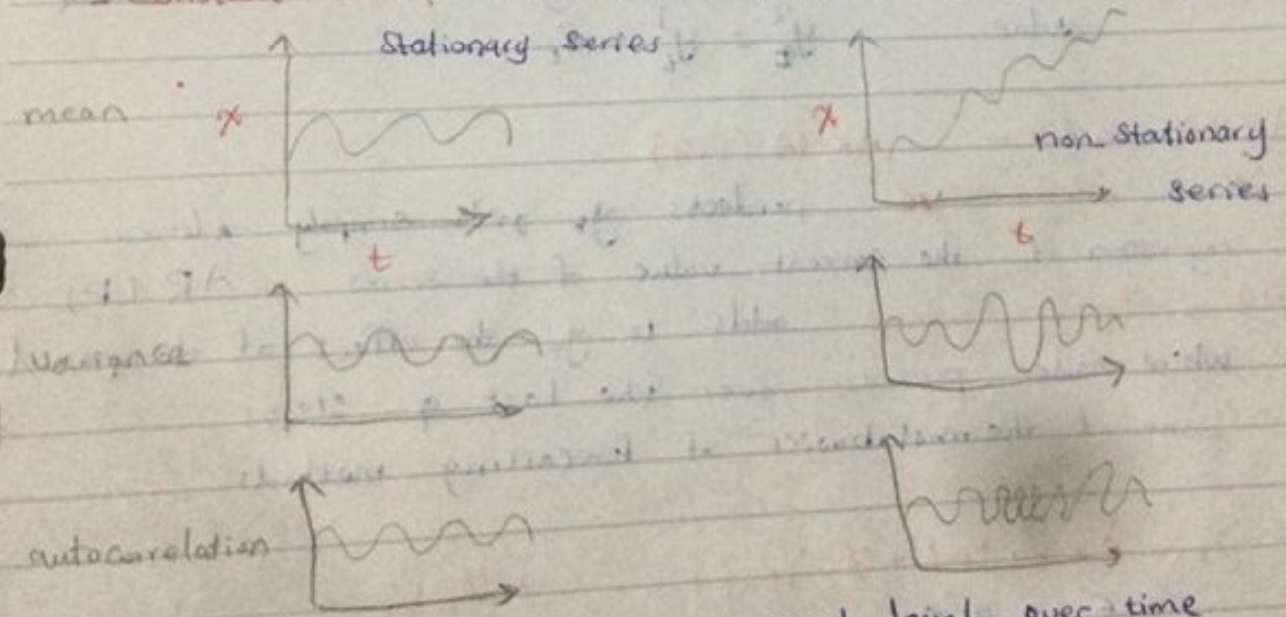
- What is Box-Jenkins Method?
  - Models historical behavior to forecast the future
  - Applies ARMA (Autoregressive + moving averages)
  - The autoregressive model specifies that the output variables depends linearly on its own previous ~~input~~ values

Input: Time series output: expected future value of Time se

$$Y_t = T_t + S_t + R_t \quad t \in 1, \ldots, n$$

Time Series = Trend term + Seasonal term + Random Component (modeled with ARMA)

- Stationary Sequences: is a random sequence in which the joint probability distribution doesn't vary overtime / or the mean, variance and autocorrelations don't change in sequence over time.
- Box-~~Jenk~~ Jenkins assumes the random component is a stationary sequence
- Constant mean or variance: the mean of the series shouldn't be a fn of time



- the series remains at a constant level over time
- If a trend exist then your data isn't stationary
- If your time series isn't stationary, you can't build time series model
- If Stationary Criterion violated first Stationarize the time series then try stochastic model to predict this time series

- In practice to obtain stationary sequence the data must be

[1] Detrending :- pre-processing step to prepare time series for analysis (remove the trend component of the series

[2] Differencing :- we try to model differences of the terms not the actual term by subtracting the observation in the current period from the previous one ; this deminates the trend if the series growing at const. rate

- If this is done only once to a series, the data is said to be ~ first differenced $X(t) - X(t-1) = ARMA(P,q)$

AR : P

I : d ARIMA → P d q

MA : q

[3] Seasonal adjustment : its done by taking several years of data, calculating average value for each month and subtracting them from the actual value $y_t^2 = y_t^1 - S_t$

- ARMA modeling ARMA (P,q)

AR → autoregressive predicts y_t , it's simply a linear regression of the current value of the series AR (P)

MA → Moving average adds to y_t the effect of dampened white noise process over the last q steps

- one of the most basics of forecasting methods

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} \quad \text{AR}$$

$$+ \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} \quad \text{MA}$$

ARIMA (P,d,q) ; adds differentiation term to ARMA

Autoregressive integrated moving average

- linear trend can be removed by d=1

- quadratic by d=2 and so on

P → no. of AR terms d → no. of differences

q → no. of MA terms

* Diagnosis : ACF & PACF

- autoCorrelations : numerical values that indicate how a data series is related to itself over time
- ACF (Auto Correlation function)
  - Correlation of the values of the time series with itself
  - Helps to determind the order „q" of MA model
- PACF (= partial ACF)
  - an auto correlation calculated after removing the linear dependence of the previous terms
  - Helps to determind the order p of an AR model

* Model Selection :

- based on the data, the data scientist select p, q, d
- use a simple model when possible (AR Model (q=0) • MA model (p=0))

- How to do a time series analysis ?

[1] Visualize the time series : Explore data to Know whether the series is stationary or not

[2] Stationarize the series : using detrending, differencing and seasonal adjustment

[3] plot ACF/PACF : to find optimal parameter p, d, q
- If both ACF/PACF decrease gradually it indicates that we need to make the series stationary and introduce a value for d

[4] Build ARIMA model : with the parameters found but we need to explore more (p, d, q) combinations

[5] Make predictions : once we have ARIMA model we are now ready to make predictions on future time po

- Reasons to choose (+) & Cautions (-) slide 28